

Elias Bound for General Distances and Stable Sets in Edge-Weighted Graphs

Marco Dalai, *Member, IEEE*

Abstract

This paper presents an extension of the Elias bound on the minimum distance of codes for discrete alphabets with general, possibly infinite-valued, distances. The bound is obtained by combining a previous extension of the Elias bound, introduced by Blahut, with an extension of a bound previously introduced by the author which builds upon ideas of Gallager, Lovász and Marton. The result can in fact be interpreted as a unification of the Elias bound and of Lovász's bound on graph (or zero-error) capacity, both being recovered as particular cases of the one presented here. Previous extensions of the Elias bound by Berlekamp, Blahut and Piret are shown to be included as particular cases of our bound. Applications to the reliability function are then discussed.

I. INTRODUCTION

A central problem in coding theory is that of determining the asymptotic performance of optimal codes when the block length is sent to infinity. Mathematically, this general problem can be formalized in different ways. An example is that of determining the asymptotic minimum distance of binary codes at a given rate. Another example is that of determining the largest possible rate for zero-error communication for a channel over which certain pairs of symbols cannot be confused. Again, another question is the determination of the asymptotic behavior of the probability of error of optimal codes at a given rate. In this paper, we consider a generalization of the first example, that is, the study of the minimum distance of codes at given rate. The proposed approach, however, borrows ideas from bounds originally developed in the context of the second example, that is bounds on the zero-error capacity of channels.

The Elias bound is certainly one of the most famous bounds on the minimum distance of codes. Originally developed for binary codes, it was later extended by Berlekamp [3], Blahut [2] and Piret [4] to more general contexts and for particular distances. In this paper, we present an extension of the bound to general, possibly infinite-valued, distances. Allowing infinite distances, we handle in a unified way bounds on the minimum distance of codes and bounds on graph (or zero-error) capacity. In fact, we propose a method which builds upon a combination of the Elias bound with the Lovász theta function to bound the minimum distance of codes even in cases where some pairs of symbols have infinite distance. The derived bound will be shown to include as special cases both the

M. Dalai is with the Department of Information Engineering, University of Brescia, Italy, e-mail: marco.dalai@unibs.it

Parts of these results were first presented in [1].

mentioned previous extensions of the Elias bound and Lovász' [5] and Marton's [6] bounds on graph/zero-error capacity.

The bound derived in this paper represents an evolution of some results presented in [7], [8]. The focus was there on the Bhattacharyya distance as a mean for bounding the reliability function of channels. In deriving the new bound, we present the result with generality for an arbitrary distance. We then discuss the particular application to the Bhattacharyya distance and to other distances that can be used to bound the reliability function.

II. NOTATION AND PROBLEM DEFINITION

A. Minimum Distance of Codes

Let \mathcal{X} be a discrete set and let d be a function $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^+ \cup \{+\infty\}$ such that for all $x, x' \in \mathcal{X}$

$$\begin{aligned} d(x, x') &\geq 0 \\ d(x, x') &= d(x', x) \\ d(x, x) &= 0. \end{aligned}$$

We will refer to such a function d as a “distance”, although as seen above we do not really require all the properties of a distance. We stress that d is allowed to take value $+\infty$ for some pairs of symbols, a case which will be of practical interest in this paper. We extend the distance to sequences of symbols defining, for $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{x}' = (x'_1, \dots, x'_n)$,

$$d(\mathbf{x}, \mathbf{x}') := \sum_{i=1}^n d(x_i, x'_i). \quad (1)$$

Note in particular that $d(\mathbf{x}, \mathbf{x}') = \infty$ iff $d(x_i, x'_i) = \infty$ for at least one i .

A code of length n is a set $\mathcal{C} = \{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ of elements in \mathcal{X}^n , that is, sequences of n symbols from \mathcal{X} called codewords. The rate of the code is $R = \log M/n$. For a given code \mathcal{C} , we define its minimum distance as

$$d_{\min}(\mathcal{C}) := \min_{\mathbf{x}, \mathbf{x}' \in \mathcal{C}, \mathbf{x} \neq \mathbf{x}'} d(\mathbf{x}, \mathbf{x}'). \quad (2)$$

For a fixed rate R and block length n we define the optimal minimum distance as

$$d(R, n) := \max_{\mathcal{C}} d_{\min}(\mathcal{C}), \quad (3)$$

where the maximum is over all codes of length n and rate at least R . Finally, for a fixed R we define the asymptotic normalized optimal minimum distance as

$$\delta^*(R) := \limsup_{n \rightarrow \infty} \frac{1}{n} d(R, n). \quad (4)$$

Similarly, for $\delta \in [0, \infty]$, we also define the maximum rate achievable by codes with minimum distance δ^* as

$$R^*(\delta) := \sup\{R : \delta^*(R) \geq \delta\}. \quad (5)$$

We stress that $\delta = \infty$ is allowed in our formulation and, consequently, the value $R^*(\infty)$ is also of importance.

In this paper, we are interested in determining upper bounds on $\delta^*(R)$ and $R^*(\delta)$. We are however also interested in bounding the minimum distance of codes with constant composition (or *type*, see [9]). Using the notation of [9], this means that there exists a distribution P such that $\mathbf{x} \in \mathcal{T}_P^n$, $\forall \mathbf{x} \in \mathcal{C}$, where \mathcal{T}_P^n is the set of sequences of length n which contain any symbol x exactly $nP(x)$ times. We call P the composition of the codewords.

We thus introduce the basic quantities that we need to consider in this case. In particular, for a valid composition P , we define

$$d(R, n, P) := \max_{\mathcal{C}} d_{\min}(\mathcal{C}), \quad (6)$$

where the maximum is over all codes of length n , rate at least R , and composition P . Similarly, for a fixed R , we define

$$\delta^*(R, P) := \limsup_{n \rightarrow \infty} \frac{1}{n} d(R, n, P_n). \quad (7)$$

with the constraint that the P_n are valid compositions which tend to P as $n \rightarrow \infty$, that is

$$\lim_{n \rightarrow \infty} P_n(x) = P(x), \quad \forall x. \quad (8)$$

B. Matrix/Graph Theoretic Formulation

Given an undirected graph G with vertex set \mathcal{V} and edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, a stable set of nodes is a set of nodes no two (distinct) of which are adjacent in G . The stability number $\alpha(G)$ is defined as the size of a largest stable set of nodes. The Shannon capacity of the graph G is defined as

$$C(G) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha(G^{\otimes n}). \quad (9)$$

where $G^{\otimes n}$ is the n -fold strong product of the graph G with itself (see [10] for details). A constant composition analogue of this quantity can be defined as follows [6], [11]. For a composition P , consider the subgraph $G^{\otimes n}(P)$ of $G^{\otimes n}$ induced by the subset of the nodes of $G^{\otimes n}$ associated to sequences of composition P . We define

$$C(G, P) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha(G^{\otimes n}(P_n)), \quad (10)$$

where the P_n are valid compositions which tend to P as $n \rightarrow \infty$.

We can extend these definitions so as to include an equivalent formulation of the minimum distance problem. We consider graphs weighted on the edges, which we identify with the matrix G of the edge weights $g(v, v')$. Here we assume that $g(v, v') \in [0, 1]$ and that $g(v, v) = 1$. We say that a set \mathcal{C} of nodes is ϵ -stable if $g(v, v') \leq \epsilon$ for $v, v' \in \mathcal{C}$ with $v \neq v'$. Then we define $\alpha(G; \epsilon)$ as the size of a largest ϵ -stable set. If we now consider the n -fold Kronecker power $G^{\otimes n}$ of the matrix G , we find that an exponential number of the off-diagonal entries approach zero exponentially fast in n . We can then define the ϵ -capacity of the graph as

$$C(G; \epsilon) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha(G^{\otimes n}; \epsilon^n). \quad (11)$$

Note that, when specialized to $\epsilon = 0$, this definition recovers equation (9) if distinct nodes are considered adjacent if and only if $g(x, x') > 0$. In this sense, $C(G; \epsilon)$ generalizes $C(G)$ in such a way that $C(G; 0) = C(G)$. Similarly,

we can define

$$C(G, P; \epsilon) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha(G^{\otimes n}(P_n); \epsilon^n) \quad (12)$$

where again the P_n are valid compositions which tend to P as $n \rightarrow \infty$.

The problem defined in the previous section can now be reformulated in this setting by considering a graph with vertex set \mathcal{X} and edge weights

$$g(x, x') := e^{-d(x, x')}, \quad (13)$$

with the convention that $e^{-\infty} = 0$. Since $d(x, x')$ is a measure of the distance between symbols, the function $g(x, x')$ represents a measure of similarity which varies from 1 to 0. Then, we can observe that we have

$$R^*(\delta) = C(G, e^{-\delta}). \quad (14)$$

We now present two examples which clarify the generality of the considered problem when we allow infinite values for the distance $d(x, x')$ and the usefulness of the graph theoretic formulation.

Example 1 (Elias Bound for Binary Codes): In this case, $\mathcal{X} = \{0, 1\}$ and d is the Hamming distance d_H defined by setting $d_H(1, 0) = 1$. Defining the binary entropy function

$$h(t) := -t \log t - (1 - t) \log(1 - t), \quad (15)$$

the Elias bound states that if

$$R = \log(2) - h(\lambda), \quad 0 \leq \lambda < 1/2 \quad (16)$$

then

$$\delta_H^*(R) \leq 2\lambda(1 - \lambda). \quad (17)$$

(see the next section for a proof).

The graph representation of the problem is obtained by using the matrix

$$G = \begin{pmatrix} 1 & e^{-1} \\ e^{-1} & 1 \end{pmatrix}. \quad (18)$$

It is easily checked that the (x, x') -entry of the matrix $G^{\otimes n}$ has value $e^{-d_H(x, x')}$.

Example 2 (Lovász' Bound on Graph Capacity [5]): Lovász' upper bound to the capacity $C(G)$ of an ordinary undirected graph G can be stated as follows. Let $\{u_x\}$ be a set of unit norm vectors in a Hilbert space. We say that $\{u_x\}$ is an orthogonal representation of the graph G if $u_x^\dagger u_{x'} = 0$ whenever x and x' are not adjacent in G . Define the quantity¹

$$\vartheta(G) := \min_{\{u_x\}, f} \max_x \log \frac{1}{|u_x^\dagger f|^2}, \quad (19)$$

where the minimum is over all orthogonal representations $\{u_x\}$ and unit norm vectors f . Then,

$$C(G) \leq \vartheta(G). \quad (20)$$

¹We point out that we use a logarithmic definition of the Lovász theta function for ease of comparison with rates and other quantities that we will need in this paper.

Note that the problem of determining the graph capacity can be stated in terms of minimum distance of codes by defining a distance $d(x, x')$ such that $d(x, x') = \infty$ if and only if x and x' are not connected in G . Then, $C(G) = R^*(\infty)$.

For the sake of simplicity, we will present our results with a focus on the minimum distance interpretation, but it is clear that an equivalent formulation of each single result can be given in terms of ϵ -capacity of weighted graphs. We will point out some of these formulations when useful.

III. PREVIOUS EXTENSIONS OF THE ELIAS BOUND

Generalizations of the Elias bound to non-binary codes have already appeared in the literature. The main contributions in this direction are those of Berlekamp [3, Ch. 13], Blahut [2], and Piret [4]. Those bounds can be considered extensions of the original Elias bound since they are based on the same basic idea. For a given code, one first identifies a subset \mathcal{T} of codewords which are all *packed* in a ball around a properly chosen fixed sequence \bar{x} . Then, the Plotkin bound is used to bound the minimum distance of the code in terms of the average distance between pairs of distinct codewords in \mathcal{T} as

$$d_{\min} \leq \frac{1}{|\mathcal{T}|(|\mathcal{T}| - 1)} \sum_{\mathbf{x}, \mathbf{x}' \in \mathcal{T}} d(\mathbf{x}, \mathbf{x}'). \quad (21)$$

The average in eq. (21) can be computed in terms of the componentwise distances as

$$\sum_{\mathbf{x}, \mathbf{x}' \in \mathcal{T}} d(\mathbf{x}, \mathbf{x}') = \sum_{\mathbf{x}, \mathbf{x}' \in \mathcal{T}} \sum_{i=1}^n d(x_i, x'_i) \quad (22)$$

$$= \sum_{i=1}^n \left(\sum_{\mathbf{x}, \mathbf{x}' \in \mathcal{T}} d(x_i, x'_i) \right). \quad (23)$$

Then, the constraints on the compositions of the sequences \mathbf{x}, \mathbf{x}' (and possibly \bar{x}) are used to derive the final bound both in the original formulation for binary codes and in slightly different ways in the more general contexts considered in [3], [2], [4].

In this Section, we first recall for the reader convenience how the approach sketched above is used in the binary case to derive the original bound stated in Example 1. Then, we discuss the extensions proposed in [3, Ch. 13], [2], [4] and provide a high level description of the bound introduced in this work.

A. Binary Codes

For any n -bit sequence \mathbf{x} , let $\mathcal{B}_w(\mathbf{x})$ be the set of n -bit binary sequences which differ from \mathbf{x} in exactly w positions. It is well known that $|\mathcal{B}_w(\mathbf{x})| = e^{n(h(w/n) + \varepsilon_n)}$, where h is the binary entropy function defined in (15), and $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$. Hence, by symmetry, any sequence $\mathbf{x}' \in \{0, 1\}^n$ is contained in $\mathcal{B}_w(\mathbf{x})$ for $e^{nh(w/n) + \varepsilon_n}$ sequences \mathbf{x} . Given any code with M codewords, the total number of codewords counted (with multiplicities) in all the sets $\mathcal{B}_w(\mathbf{x})$, as \mathbf{x} runs over $\{0, 1\}^n$, is thus $Me^{nh(w/n) + \varepsilon_n}$. Since there are 2^n sets $\mathcal{B}_w(\mathbf{x})$, at least one of them, say $\mathcal{B}_w(\bar{x})$, contains $Me^{n(h(w/n) + \varepsilon_n - \log(2))}$ codewords. Note that these codewords are all packed around the sequence \bar{x} since they all differ from it in exactly w positions. We can thus use equation (21) with the choice

$\mathcal{T} = \mathcal{B}_w(\bar{x}) \cap \mathcal{C}$. Note that $|\mathcal{T}| = e^{n(R+h(w/n)-\log(2)+\varepsilon_n)}$. If we call v_i the number of codewords in \mathcal{T} which differ from \bar{x} in the coordinate i , it can be checked that equation (23) can be continued as

$$\sum_{i=1}^n \left(\sum_{x, x' \in \mathcal{T}} d(x_i, x'_i) \right) = \sum_{i=1}^n 2v_i(1 - v_i). \quad (24)$$

But, from the definition of \mathcal{T} we must have

$$\sum_{i=1}^n v_i = w|\mathcal{T}|. \quad (25)$$

Defining $\lambda_i = v_i/|\mathcal{T}|$, we thus have from (21) and (24)

$$d_{\min} \leq \frac{|\mathcal{T}|}{|\mathcal{T}| - 1} \sum_{i=1}^n 2\lambda_i(1 - \lambda_i), \quad (26)$$

where the λ_i satisfy

$$\frac{1}{n} \sum_{i=1}^n \lambda_i = \frac{w}{n}. \quad (27)$$

Since the function $\lambda(1 - \lambda)$ is concave², we can apply the Jensen inequality to equation (26) to obtain

$$d_{\min} \leq 2n \frac{|\mathcal{T}|}{|\mathcal{T}| - 1} \left(\frac{w}{n} \right) \left(1 - \frac{w}{n} \right). \quad (28)$$

When n goes to infinity, $|\mathcal{T}| = e^{n(R+h(w/n)-\log(2)+\varepsilon_n)}$ goes to infinity if w is chosen as a function of n in such a way that w/n tends to a limit λ such that $R > \log(2) - h(\lambda)$. We then obtain the bound

$$\delta^*(R) \leq 2\lambda(1 - \lambda) \quad (29)$$

valid for $R > \log(2) - h(\lambda)$. By continuity of $\lambda(1 - \lambda)$, we can extend the bound to the case $R = \log(2) - h(\lambda)$.

B. Berlekamp's Bound

Berlekamp considers two possible choices of the distance d . Assuming \mathcal{X} is the set $\mathbb{Z}_K = \{0, \dots, K - 1\}$, endowed with the usual sum mod K , Berlekamp considers the Hamming distance defined by

$$d_H(x, x') = \begin{cases} 0 & \text{if } x = x' \\ 1 & \text{if } x \neq x', \end{cases} \quad (30)$$

and the Lee distance defined by

$$d_L(x, x') = \min(x - x', x' - x). \quad (31)$$

Note that in all cases the distance is finite. The bound in [3] is stated as follows. Let $V_d^{(n)}$ be the number of sequences of length n with distance at most d from the sequence $\mathbf{0} = (0, 0, \dots, 0)$. Let then

$$t(R, n) = \frac{1}{n} \min\{d : V_d^{(n)} e^{nR} \geq K^n\} \quad (32)$$

and

$$t(R) = \liminf_{n \rightarrow \infty} t(R, n). \quad (33)$$

²This concavity argument is fundamental to all versions of the Elias bound, see Section III-E.

Then

$$\delta^*(R) \leq t(R) \left(2 - \frac{t(R)}{d(U)} \right), \quad (34)$$

where $d(U)$ is the average distance of the points in \mathcal{X} from 0 under a uniform distribution

$$d(U) = \sum_x \frac{1}{K} d(0, x). \quad (35)$$

It is useful to rewrite Berlekamp's bound in a more convenient way. Note that all sequences with a given composition Q are at the same distance from $\mathbf{0}$, which is given by the expression

$$d(\mathbf{x}, \mathbf{0}) = n \sum_x Q(x) d(0, x). \quad (36)$$

Since the number of different compositions of sequences in \mathcal{X}^n is polynomial in n , the exponential growth of the quantity $V_d^{(n)}$ is determined by the largest composition class at distance at most d from $\mathbf{0}$. To the first order in the exponent, there are $e^{nH(Q)}$ sequences of composition Q , where $H(\cdot)$ is the entropy of a distribution. Hence, we will have

$$V_d^{(n)} = e^{n(H(Q) + o(1))}, \quad (37)$$

where Q maximizes $H(Q)$ over the set of compositions satisfying

$$\sum_x Q(x) d(0, x) \leq d. \quad (38)$$

Taking the limit as $n \rightarrow \infty$, this implies that

$$t(R) = \min_{Q \in \mathcal{Q}(R)} \sum_x Q(x) d(0, x) \quad (39)$$

where

$$\mathcal{Q}(R) = \{Q : R + H(Q) \geq \log K\}. \quad (40)$$

Since $\mathcal{Q}(R)$ is a convex domain and the objective function is linear, it can be verified with the use of Lagrange multipliers that the optimal Q^* satisfies

$$Q^*(x) = \eta e^{-\mu d(0, x)} \quad (41)$$

where η, μ are parameters chosen so as to satisfy the constraints on Q with equality.

An important remark about this bound is that it relies only on two properties of the considered distances. The first property is that the distances are circularly symmetric, that is

$$d(x, x') = d(0, x - x') \quad (42)$$

and, hence, circular permutations of the elements in \mathcal{X} do not change the cross-distances. The second property is that the quadratic form

$$\sum_{x, x'} Q(x) Q(x') d(x, x') \quad (43)$$

is a concave function over the simplex of distributions Q (see [3, Sec. 13.4 and eqs. (13.63)-(13.66)]). This requirement is essentially needed in order to apply the Jensen inequality as we did to move from equation (26) to

equation (28) in the binary case. As we will see below, this is a recurrent requirement in all versions of the Elias bound and it is essentially due to the use of the Plotkin bound.

C. Blahut's Bound

Blahut develops his bound on the minimum distance as a tool for bounding the reliability function of discrete memoryless channels. Let $W_x(y)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, be the transition probabilities of a discrete memoryless channel W with input alphabet \mathcal{X} and output alphabet \mathcal{Y} . Blahut considers the Bhattacharyya distance between symbols, defined by

$$d_B(x, x') = -\log \sum_y \sqrt{W_x(y)W_{x'}(y)}, \quad (44)$$

under the assumption that $d_B(x, x')$ is finite, which means that the channel W has no zero-error capacity [12]. His bound, which applies to a particular subset of channels to be described below, in the most general form states that

$$\delta_B^*(R, P) \leq \min_{V \in \mathcal{V}(P)} \sum_{x, x_1, x_2} P(x) V_x(x_1) V_x(x_2) d(x_1, x_2) \quad (45)$$

where $V = \{V_x(x')\}_{x, x' \in \mathcal{X}}$ is a stochastic matrix running over the set

$$\mathcal{V}(P) = \{V : PV = P, I(P, V) \leq R\}. \quad (46)$$

Here, $I(P, V)$ is the mutual information with the notation of [9], and $PV = P$ means

$$\sum_x P(x) V_x(x') = P(x'). \quad (47)$$

After optimization of the composition we then have

$$\delta_B^*(R) \leq \max_P \min_{V \in \mathcal{V}(P)} \sum_{x, x_1, x_2} P(x) V_x(x_1) V_x(x_2) d(x_1, x_2). \quad (48)$$

The particular class of channels for which this bound is derived by Blahut is that of the so-called non-negative definite channels studied by Jelinek [13], which are channels for which the matrix $G(\rho)$ with elements (see notations (13) and (44))

$$g(x, x')^{1/\rho} = \left(\sum_y \sqrt{W_x(y)W_{x'}(y)} \right)^{1/\rho} \quad (49)$$

is positive semidefinite for all $\rho \geq 1$. As we will discuss later (see Lemma 1 below), this property implies that the quadratic form in equation (43) is concave³ in the distribution Q , and this is essentially the only property of the distance used in Blahut's derivation. Note in particular that, contrarily to Berlekamp's bound, there is no algebraic structure in the alphabet and no circular symmetry of $d_B(x, x')$ in the sense of (42), which is the reason for the more complicated resulting expression in the bound.

³Compare also with [2, Lemma 5]; note that there is a sign error in the derivation of the proof and, thus, it is actually proved that the quadratic form is concave and not convex as stated in the lemma.

D. Piret's Bound

Piret considers the case where the elements of \mathcal{X} are uniformly spaced points on the unit circle in the plane and the distance is the squared euclidean distance. Note that this is a case where the distance is explicitly assumed to be a circularly symmetric squared euclidean distance. Piret's bound follows an approach which is similar to Blahut's and can be stated as follows. Let again

$$\mathcal{Q}(R) = \{Q : R + H(Q) \geq \log K\}. \quad (50)$$

Then, for any $Q \in \mathcal{Q}(R)$ we have

$$\delta^*(R) \leq \sum_{x, x'} Q(x)Q(x')d(x, x'). \quad (51)$$

Again, the only properties used by Piret are the circular symmetry of the distance and the concavity of the quadratic form (see [4, Lemma 4.2]). The latter probably leads Piret to observe that it is difficult to determine the distribution $Q \in \mathcal{Q}(R)$ which gives the best bound, and he thus suggests to search heuristically for a Q which gives good bounds. Interestingly, even if apparently unaware of Berlekamp's approach, he conjectures that the optimal Q is the same Q^* defined in Section III-B which minimizes (39).

E. Discussion: Squared Euclidean Distances

A detailed discussion of the connections between Berlekamp's, Blahut's and Piret's bounds is of interest and, to the best of the author's knowledge, has not been reported previously in the literature. We postpone such an analysis to a later section where we show that all three bounds are included as special cases in our bound. It will turn out that our bound includes Blahut's one which in turn includes Piret's bound, which finally implies Berlekamp's one.

For the moment, instead, we discuss the fundamental common feature that has already emerged in the presentation of the bounds, which is the fact that they all rely on the concavity of the quadratic form (43) on the simplex of probability distributions. In order to better investigate this property, we need the following lemma, whose proof is given in the Appendix.

Lemma 1: Let \mathcal{X} be a finite alphabet and $d(x, x')$ a distance on \mathcal{X} as defined in Section II. If d is finite, then the following four statements are equivalent

- (a) The $|\mathcal{X}| \times |\mathcal{X}|$ matrix $G(\rho)$ with (x, x') element $e^{-d(x, x')/\rho}$ is positive semidefinite for all $\rho > 0$.
- (b) We have

$$\sum_{x, x'} c(x)c(x')d(x, x') \leq 0, \quad \text{provided } \sum_{x \in \mathcal{X}} c(x) = 0. \quad (52)$$

- (c) The quadratic form

$$\sum_{x, x'} Q(x)Q(x')d(x, x') \quad (53)$$

is concave function of Q on the simplex of probability distributions.

- (d) $(\mathcal{X}, d^{1/2})$ is a metric space which can be embedded into a euclidean space, that is, d is a squared euclidean distance.

According to the Lemma above, for finite distances the quadratic form (43) is concave if and only if the distance d is a squared euclidean distance. That is, there exist points v_x in a euclidean space such that $d(x, x') = \|v_x - v_{x'}\|_2^2$. This observation, which does not seem to have been made before in this context, automatically implies that not only the distance used by Piret, but also those used by Berlekamp and Blahut are squared euclidean distances. Conversely, for the same reason, since the concavity of (43) is the only used property, Berlekamp's bound applies to all circularly symmetric squared euclidean distances and Blahut's bound applies to all squared euclidean distances.

We note here that the Hamming distance is trivially representable as a squared euclidean distance in $\mathbb{R}^{|\mathcal{X}|}$ using mutually orthogonal vectors. For the Lee distance, simple embeddings can be found by considering the case of even or odd cardinality separately⁴, as shown in Fig. 1.

For the Bhattacharyya distance, finally, we observe that, as already mentioned, Blahut develops his bound only for the class of non-negative channels studied by Jelinek with the additional assumption that they have no zero error capacity. For these channels, the Bhattacharyya distance is finite and satisfies by definition⁵ the condition (a) of Lemma 1. It was already observed by Jelinek that the equivalence of conditions (a) and (b) in Lemma 1 was a known fact among algebraists (see his comments to [13, Th. 2]). However, what was apparently not noticed before in the information theory community, is the equivalence of these conditions with condition (d) of Lemma 1. In particular, this implies that for Jelinek's channels the Bhattacharyya distance is a squared euclidean distance. We observe that Jelinek also reports in his Lemma 1 another "test" for the condition (b) in our Lemma 1 to be satisfied, namely - using our notation - that the matrix \tilde{D} with elements

$$\tilde{d}(x_1, x_2) = -d(x_1, x_2) + \frac{1}{|\mathcal{X}|} \sum_{x'} d(x_1, x') + \frac{1}{|\mathcal{X}|} \sum_{x'} d(x', x_1) - \frac{1}{|\mathcal{X}|^2} \sum_{x', x''} d(x', x'') \quad (54)$$

is positive semidefinite (see the proof of Lemma 1 in Appendix). What is curious is that this property is reported by Jelinek but has apparently no importance in his paper, and even if already known in the algebraic community (see [14]) it was probably not much used outside, at that time. On the contrary, equation (54) is by now a fairly well known equation in kernel based learning theory since it represents the condition for a set of points with given cross-distances to be embeddable in a euclidean space. The requirement that the matrix \tilde{D} be positive semidefinite implies that it is the Gram matrix of a set of vectors $\{\tilde{v}_x\}$ and it turns out that the vectors $u_x = v_x/\sqrt{2}$ satisfy $\|u_x - u_{x'}\|^2 = d(x, x')$. Hence, Jelinek's channels are precisely those for which the Bhattacharyya distance is a squared euclidean distance. We point out that Jelinek also considers non-negative channels with a zero-error capacity. In this case, the set \mathcal{X} can be partitioned in subsets such that the Bhattacharyya distance is a squared euclidean distance within each subset and it is infinite between symbols from different subsets. So, it can still be interpreted as an euclidean distance if we allow these subsets to be infinitely far apart in the space.

⁴I am indebted to an anonymous correspondent which goes under the nickname "El Filibustero" for pointing out these simple embeddings.

⁵Jelinek only asks the condition for $\rho \geq 1$, but it is not difficult to show that this actually implies it for all $\rho > 0$, since the element-wise product of positive semidefinite matrices is positive semidefinite. See also the proof of Lemma 1.

$$\begin{array}{rcc}
& & \begin{matrix} 0 & \alpha & \alpha & 0 & 0 \\ 0 & \alpha & \alpha & \alpha & 0 \\ 0 & 0 & \alpha & \alpha & 0 \\ 0 & 0 & \alpha & \alpha & \alpha \\ 0 & 0 & 0 & \alpha & \alpha \end{matrix} \\
|\mathcal{X}| = 5 : & & \\
& & \\
& & \begin{matrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{matrix} \\
|\mathcal{X}| = 6 : & &
\end{array}$$

Fig. 1. Example of (squared) euclidean embedding of the Lee distance (see footnote 4). The columns of the above matrices, where $\alpha = 1/\sqrt{2}$, as points in \mathbb{R}^5 and \mathbb{R}^3 respectively, have squared euclidean distances which match respectively the Lee distance over \mathbb{Z}_5 and \mathbb{Z}_6 . Extensions to arbitrary values of $|\mathcal{X}|$ is obvious.

F. Our Bound

The extension of the Elias bound that we propose is primarily motivated by the need to deal with infinite distances. This is a necessary step when considering channels with a zero-error capacity for which some pairs of symbols cannot be confused. Then, any reasonable distance must take infinite value for non-confusable symbols. This is in fact the case for example with the Bhattacharyya distance.

Our approach is based on a variation of the Plotkin step. In a nutshell, since we want to cope with infinite distances, rather than averaging the pairwise distances $d(\mathbf{x}, \mathbf{x}')$, we average an exponential function of those distances. In particular, we use an approach which in a sense corresponds to substituting equation (21) with

$$d_{\min} \leq -\rho \log \left(\max_{\mathbf{x} \in \mathcal{T}} \frac{1}{(|\mathcal{T}| - 1)} \sum_{\mathbf{x}' \in \mathcal{T} \setminus \{\mathbf{x}\}} e^{-d(\mathbf{x}, \mathbf{x}')/\rho} \right). \quad (55)$$

There is a drawback of course, in that the derivation of the bound must now follow a different route, since it is no longer possible to use eq. (23). We approach the problem by proposing an extension of the umbrella bound originally introduced in [7]. That bound can in fact be interpreted as a variation of the Plotkin bound (21) in the form of equation (55), when there is no constraint on the composition of the codewords \mathbf{x}, \mathbf{x}' . Here, we propose an extension of the method that allows us to handle composition constraints as is usually done with equation (23).

IV. EXTENSION OF THE ELIAS BOUND

In this section we present our extension of the Elias bound to the case of general, possibly infinite-valued, distances d . The extension combines the two basic ideas used in the Elias bound and in Lovász' bound on the zero error capacity. A first step in this unification consists in extending the ordinary Lovász bound to the case of graphs weighted on the edges as presented in Section II-B. This was already done in previous works [7], [8], although the presentation was given for the particular case where the edge weights are related to the Bhattacharyya distances

between input symbols of a discrete memoryless channel. We first review that original extension of the ϑ function in the notation of the present paper, and then we present the additional required extensions and the combination with Elias's bounding procedure.

For the sake of simplicity and for coherence with the literature on Elias' bounds, we develop our procedure with a focus on the rate-distance relation in terms of $\delta^*(R)$ and $R^*(\delta)$ functions. As explained before, however, it is clear that the procedure could be stated solely in terms of weighted graphs and their ϵ -capacities by considering edge weights $g(x, x')$ associated to the given distance d according to $g(x, x') = e^{-d(x, x')}$. We will only briefly mention the analogous expressions in that context and give the main derivation in terms of distances.

A. The $\vartheta(\rho)$ function

The function $\vartheta(\rho)$ introduced in [7] can be defined as follows. Given the set \mathcal{X} and distance d , for a fixed⁶ $\rho > 0$, an *orthonormal representation of degree ρ* of our distance is a set of unit norm vectors $\{u_x\}$ in any Hilbert space such that $|u_x^\dagger u_{x'}| \leq e^{-d(x, x')/\rho}$. Call $\Gamma(\rho)$ the non-empty set of all possible such representations

$$\Gamma(\rho) = \left\{ \{u_x\} : |u_x^\dagger u_{x'}| \leq e^{-d(x, x')/\rho} \right\}, \quad \rho > 0. \quad (56)$$

The *value* of an orthonormal representation is the quantity

$$V(\{u_x\}) = \min_f \max_x \log \frac{1}{|u_x^\dagger f|^2}, \quad (57)$$

where the minimum is over all unit norm vectors f . The optimal choice of the vector f is called the *handle* of the representation. The function $\vartheta(\rho)$ is defined as the minimum value over all representations of degree ρ , that is,

$$\vartheta(\rho) = \min_{\{u_x\} \in \Gamma(\rho)} V(\{u_x\}). \quad (58)$$

The result presented in [7] can be stated as (a slightly different form of) the following theorem.

Theorem 1: For a code \mathcal{C} of block-length n with M codewords and any $\rho > 0$, we have

$$d_{\min}(\mathcal{C}) \leq -\rho \log \left(\frac{M e^{-n\vartheta(\rho)} - 1}{M - 1} \right).$$

This result is essentially based on the following Lemma, which we will also need in this paper and that we prove here for convenience.

Lemma 2: Let v_1, \dots, v_M and w be unit norm vectors such that $|v_i^\dagger w|^2 \geq c > 0$ for all i . Then

$$\max_{i \neq j} |v_i^\dagger v_j| \geq \frac{Mc - 1}{M - 1}.$$

Proof: Let Φ be a matrix whose i -th column is v_i . Then, direct computation shows that

$$w^\dagger \Phi \Phi^\dagger w \geq Mc.$$

⁶Only values of $\rho \geq 1$ were considered in [7], [8]. Here, the way we use $\vartheta(\rho)$ (compare Theorem 1 below with [7, Th. 1]) allows arbitrary positive values of ρ .

Since w is a unit norm vector, $\lambda_{\max}(\Phi\Phi^\dagger) \geq Mc$, where λ_{\max} is the largest eigenvalue. This also implies $\lambda_{\max}(\Phi^\dagger\Phi) \geq Mc$. For a matrix A with elements $a(i, j)$, it is known that

$$\lambda_{\max}(A) \leq \max_i \sum_j |a(i, j)|. \quad (59)$$

Applying this to $A = \Phi^\dagger\Phi$ we obtain

$$\begin{aligned} Mc &\leq \lambda_{\max}(\Phi^\dagger\Phi) \\ &\leq \max_i \sum_j |v_i^\dagger v_j| \\ &\leq 1 + (M-1) \max_{i \neq j} |v_i^\dagger v_j| \end{aligned}$$

which implies the statement of the lemma. \blacksquare

Given a representation $\{u_x\}$ with handle f achieving $\vartheta(\rho)$, we can associate to a sequence $\mathbf{x} = (x_1, \dots, x_n)$ the vector

$$\mathbf{u}_{\mathbf{x}} = u_{x_1} \otimes \dots \otimes u_{x_n}. \quad (60)$$

Setting $\mathbf{f} = f^{\otimes n}$, we find

$$|\mathbf{u}_{\mathbf{x}}^\dagger \mathbf{f}|^2 = \prod_{i=1}^n |u_{x_i}^\dagger f|^2 \quad (61)$$

$$\geq e^{-n\vartheta(\rho)}. \quad (62)$$

Hence, for a code $\mathcal{C} = \{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ Lemma 2, used with the vectors $\mathbf{u}_{\mathbf{x}_i}$ in place of the v_i 's and \mathbf{f} in place of w , implies that

$$\max_{m \neq m'} |\mathbf{u}_{\mathbf{x}_m}^\dagger \mathbf{u}_{\mathbf{x}_{m'}}| \geq \frac{Me^{-n\vartheta(\rho)} - 1}{M-1}. \quad (63)$$

On the other hand, we have

$$|\mathbf{u}_{\mathbf{x}}^\dagger \mathbf{u}_{\mathbf{x}'}| = \prod_{i=1}^n |u_{x_i}^\dagger u_{x'_i}| \quad (64)$$

$$\leq \prod_{i=1}^n e^{-d(x_i, x'_i)/\rho} \quad (65)$$

$$= e^{-d(\mathbf{x}, \mathbf{x}')/\rho} \quad (66)$$

and, hence,

$$d_{\min}(\mathcal{C}) \leq -\rho \log \left(\max_{m \neq m'} |\mathbf{u}_{\mathbf{x}_m}^\dagger \mathbf{u}_{\mathbf{x}_{m'}}| \right). \quad (67)$$

Combining equation (63) and (67) we obtain Theorem 1.

When considering the asymptotic regime $n \rightarrow \infty$, Theorem 1 implies the following bound on $\delta^*(R)$ (cf. [7], [8]).

Theorem 2: For any value of $\rho > 0$,

$$\text{if } R > \vartheta(\rho), \text{ then } \delta^*(R) \leq \rho\vartheta(\rho). \quad (68)$$

An equivalent formulation of these results can be stated in terms of a weighted graph G by simply letting the graph edge weights $g(x, x')$ play the same role of $e^{-d(x, x')}$ in the definition of the set $\Gamma(\rho)$ in equation (56). Here, we strengthen the notation writing $\vartheta(G, \rho)$ for clarity. Then, the results discussed before can be presented using the definition of ϵ -stable sets as follows.

Theorem 3: For a weighted graph G , $\epsilon \in [0, 1)$, and any $\rho \geq 0$, we have the bound

$$\alpha(G; \epsilon) \leq \frac{1 - \epsilon^{1/\rho}}{e^{-\vartheta(G, \rho)} + \epsilon^{1/\rho}} \quad (69)$$

Equations (62) and (66) then essentially imply that $\vartheta(G^{\otimes n}, \rho) \leq n\vartheta(G, \rho)$. So, when used for the graph $G^{\otimes n}$, Theorem 3 says that

$$\alpha(G^{\otimes n}; \epsilon^n) \leq \frac{1 - \epsilon^{n/\rho}}{e^{-n\vartheta(G, \rho)} + \epsilon^{n/\rho}}. \quad (70)$$

In the limit of $n \rightarrow \infty$, the equivalent of Theorem 2 is as follows.

Theorem 4: For a weighted graph G ,

$$\text{if } \epsilon < e^{-\rho\vartheta(G)} \text{ then } C(G; \epsilon) \leq \vartheta(G, \rho). \quad (71)$$

Remark 1: Note that the standard Lovász bound on the zero-error capacity is obtained by setting $\epsilon = 0$, which allows us to use $\rho \rightarrow \infty$ thus recovering the bound $C(G) \leq \vartheta(G)$.

Remark 2: In general, with a procedure similar to equations (62) and (66), used for moving from equation (69) to (70), we can deduce that for two graphs G_1 and G_2 ,

$$\vartheta(G_1 \otimes G_2, \rho) \leq \vartheta(G_1, \rho) + \vartheta(G_2, \rho). \quad (72)$$

For the ordinary ϑ function, we know that the equivalent expression holds with equality [5]. We have not yet investigated whether equality holds also with our extended version, but this will not be needed in the present paper.

B. Constant Composition Codes

The first step that we need to consider, for the development of a bound along the Elias scheme, is the extension of Theorem 1 to the case of codes with a constant composition. Hence, we will first modify our previous approach to bound $\delta^*(R, P)$. Note that the main property of the function $\vartheta(\rho)$ that we used is the property expressed in equation (62). There we really see the reason for the definition of $\vartheta(\rho)$. We built a set of vectors $\{u_x\}$ associated to symbols, and an auxiliary vector f such that f is “close” to all possible u_x . This in turn implies that the vector f is close to any vector u_x associated with any sequence x , no matter what the composition of x is. If we are interested in sequences x with a particular composition, however, it can be preferable to pick f so that $|u_x^\dagger f|$ is larger for the symbols x which are used more frequently in the sequence. This leads to a variation of $\vartheta(\rho)$ which is the analogue of the variation of the Lovász theta function introduced by Marton in [6] (and hence a generalization of the latter).

For a distribution P and for $\rho > 0$, we define

$$\vartheta(\rho, P) = \min_{\{u_x\} \in \Gamma(\rho), f} \sum_x P(x) \log \frac{1}{|u_x^\dagger f|^2}. \quad (73)$$

With this definition, if \mathbf{x} is a sequence with composition P , and $\{u_x\}$ is a representation with handle f achieving $\vartheta(\rho, P)$, we have

$$|\mathbf{u}_x^\dagger \mathbf{f}|^2 = \prod_{i=1}^n |u_{x_i}^\dagger f|^2 \quad (74)$$

$$= \prod_x |u_x^\dagger f|^{2nP(x)} \quad (75)$$

$$= e^{n \sum_x P(x) \log |u_x^\dagger f|^2} \quad (76)$$

$$= e^{-n\vartheta(\rho, P)}. \quad (77)$$

Consider now a code \mathcal{C} with M codewords $\mathbf{x}_1, \dots, \mathbf{x}_M$ of composition P . If we now apply again Lemma 2 to the vectors $\mathbf{u}_{\mathbf{x}_i}$ we conclude that equation (63) is simply replaced by

$$\max_{m \neq m'} |\mathbf{u}_{\mathbf{x}_m}^\dagger \mathbf{u}_{\mathbf{x}_{m'}}| \geq \frac{Me^{-n\vartheta(\rho, P)} - 1}{M - 1}. \quad (78)$$

Letting again $n \rightarrow \infty$, and using equation (67), we have the following result.

Theorem 5: For any $\rho > 0$,

$$\text{if } R > \vartheta(\rho, P), \text{ then } \delta^*(R, P) \leq \rho\vartheta(\rho, P). \quad (79)$$

Remark 3: It is obvious from the definitions that $\vartheta(\rho, P) \leq \vartheta(\rho)$ and, hence, $\max_P \vartheta(\rho, P) \leq \vartheta(\rho)$. This implies that, even after optimization of the distribution P , the bound derived here is at least as good as the one that we can derive from Theorem 1. When $\rho \rightarrow \infty$, it can be proved that in fact the equality $\max_P \vartheta(\infty, P) = \vartheta(\infty)$ holds [15]. We have not yet investigated if equality holds in general, but this will not be needed in this paper.

In the graph theory language, this result can be restated as a generalization of Theorem 4 to the case of constant composition codes or, using the nomenclature of Marton [6], to the case of probabilistic graphs. Note again that, as for Remark 1, Marton's result is obtained by setting $\epsilon = 0$ and letting $\rho \rightarrow \infty$.

C. The Elias Bound

We now extend further the definition of ϑ in order to apply the scheme developed by Blahut as a generalization of the Elias bound. What we need now is to extend the definition of $\vartheta(\rho, P)$ to deal with stochastic matrices. Given a set \mathcal{A} , a distribution F on \mathcal{A} , and a $|\mathcal{A}| \times |\mathcal{X}|$ stochastic matrix $V = \{V_a(x)\}$, $a \in \mathcal{A}, x \in \mathcal{X}$, we define

$$\vartheta(\rho, V|F) = \sum_a F(a) \vartheta(\rho, V_a) \quad (80)$$

$$= \min_{a, x} \sum_a F(a) V_a(x) \log \frac{1}{|u_{a,x}^\dagger f_a|^2} \quad (81)$$

where the minimum is over all *sequences* of representations $\{u_{a,1}, \dots, u_{a,|\mathcal{X}|}\} \in \Gamma(\rho)$, $a \in \mathcal{A}$ (one representation for each a) and over all sets of unit norm vectors $\{f_a\}$, $a \in \mathcal{A}$ (a different handle for each a).

Consider now the set of optimal representations and optimal handles which achieve $\vartheta(\rho, V|F)$. Let $\mathbf{a} = (a_1, a_2, \dots, a_n)$ be a sequence with composition F and define

$$\mathbf{f} = f_{a_1} \otimes f_{a_2} \cdots \otimes f_{a_n}. \quad (82)$$

Assume a sequence $\mathbf{x} = (x_1, x_2, \dots, x_n)$ has a conditional composition V given the sequence \mathbf{a} , which means that any symbol x appears in \mathbf{x} in exactly a fraction $V_a(x)$ of the $nF(a)$ positions in which a appears in \mathbf{a} , for any a . Consider the vector

$$\mathbf{u}_{\mathbf{x}} = u_{a_1, x_1} \otimes u_{a_2, x_2} \cdots \otimes u_{a_n, x_n} \quad (83)$$

Then, we have

$$|\mathbf{u}_{\mathbf{x}}^\dagger \mathbf{f}|^2 = \prod_{i=1}^n |u_{a_i, x_i}^\dagger f_{a_i}|^2 \quad (84)$$

$$= \prod_{a, x} |u_{a, x}^\dagger f_a|^{2nF(a)V_a(x)} \quad (85)$$

$$= e^{n \sum_{a, x} F(a)V_a(x) \log |u_{a, x}^\dagger f_a|^2} \quad (86)$$

$$= e^{-n\vartheta(\rho, V|F)}. \quad (87)$$

Applying again Lemma 2 as we did in our previous bounds, if we have a set of M codewords all with a conditional composition V from a fixed sequence \mathbf{a} with composition F , then

$$\max_{m \neq m'} |\mathbf{u}_{\mathbf{x}_m}^\dagger \mathbf{u}_{\mathbf{x}_{m'}}| \geq \frac{Me^{-n\vartheta(\rho, V|F)} - 1}{M - 1}. \quad (88)$$

In order to use this inequality for a given code, it is now necessary to consider the possible joint compositions of a subset of codewords with some given fixed auxiliary sequence $\mathbf{a} \in \mathcal{A}^n$. We need the following lemma, where we use the notation of [9] for types \mathbb{T}_P^n and V -shells $\mathbb{T}_V^n(\cdot)$.

Lemma 3: Let \mathcal{C} be a constant composition code with $\mathbf{x} \in \mathbb{T}_P^n$, $\forall \mathbf{x} \in \mathcal{C}$, and $|\mathcal{C}| = M$. Let \hat{V} be a conditional composition for sequences on a set \mathcal{A}^n given $\mathbf{x} \in \mathbb{T}_P^n$ (that is $nP(x)\hat{V}_x(a)$ is an integer) and let $F = P\hat{V}$. Then, there is a subset \mathcal{T} of at least $|\mathcal{T}| = Me^{-n(I(P, \hat{V}) + o(1))}$ codewords which all have joint composition $P \times \hat{V}$ with a fixed sequence $\mathbf{a} \in \mathbb{T}_F^n$.

Proof: The Lemma is a standard covering argument, and it is essentially a slight generalization of the argument used in [2, Th. 8]. It is well known that, for $\mathbf{x} \in \mathbb{T}_P^n$, $|\mathbb{T}_{\hat{V}}^n(\mathbf{x})| = e^{n(H(\hat{V}|P) + o(1))}$. On the other hand, if $\mathbf{a} \in \mathbb{T}_V^n(\mathbf{x})$ then $\mathbf{a} \in \mathbb{T}_F^n$, and $|\mathbb{T}_F^n| = e^{n(H(F) + o(1))}$. Hence, since $|\mathcal{C}| = M$, at least one sequence \mathbf{a} is contained in $\mathbb{T}_V^n(\mathbf{x})$ for at least $Me^{n(H(\hat{V}|P) - H(F) + o(1))} = Me^{-n(I(P, \hat{V}) + o(1))}$ codewords \mathbf{x} . ■

We can now apply the bound of equation (88) for the subset \mathcal{T} of codewords determined in Lemma 3. Let $V_a(x) = P(x)\hat{V}_x(a)/F(a)$ be the conditional composition of these codewords given the sequence \mathbf{a} . For coherence with our notation, it will be useful to express all quantities in terms of F and V rather than P and \hat{V} . From equation (88) used with the set \mathcal{T} we obtain

$$\max_{m \neq m'} |\mathbf{u}_{\mathbf{x}_m}^\dagger \mathbf{u}_{\mathbf{x}_{m'}}| \geq \frac{Me^{-n(I(F, V) + \vartheta(\rho, V|F) + o(1))} - 1}{Me^{-n(I(F, V) + o(1))} - 1}. \quad (89)$$

Asymptotically as $n \rightarrow \infty$, if the rate R is larger than $I(F, V) + \vartheta(\rho, V|F)$, both numerator and denominator in the right hand side of the above equation grow exponentially in n and their ratio is asymptotic to $e^{-n\vartheta(\rho, V|F)}$.

Using again equation (66) we have

$$\frac{1}{n} d_{\min}(\mathbf{C}) \leq \rho \vartheta(\rho, V|F) + o(1). \quad (90)$$

For fixed n , the choice of F and V is constrained to satisfy the usual type constraints, but asymptotically as $n \rightarrow \infty$ these constraints can be neglected. As a consequence, we have the following theorem.

Theorem 6: For given R, P and $\rho > 0$, let F be a distribution on a set \mathcal{A} and V be a $|\mathcal{A}| \times |\mathcal{X}|$ stochastic matrix such that $FV = P$. Then,

$$\text{if } R > I(F, V) + \vartheta(\rho, V|F), \quad \text{then } \delta^*(R, P) \leq \rho \vartheta(\rho, V|F). \quad (91)$$

Remark 4: We observe that with the choice $\mathcal{A} = \{a\}$ and $V_a(x) = P(x)$ we have $FV = P$, $I(F, V) = 0$ and $\vartheta(\rho, V|F) = \vartheta(\rho, P)$. Hence, if $R > \vartheta(\rho, P)$ for a given ρ , this particular choice gives the same bound of Theorem 5, which is thus included as a particular case in Theorem 6.

In the language of graph theory the result reads as follows.

Theorem 7: Under the same conditions of Theorem 6, for a weighted Graph G we have

$$\text{if } \epsilon < e^{-\rho \vartheta(\rho, V|F)}, \quad \text{then } C(G, P; \epsilon) \leq I(F, V) + \vartheta(\rho, V|F). \quad (92)$$

V. ANALYSIS OF THE BOUND

The evaluation of the bound presented in Theorem 6 is not simple in the general case. A complete theoretical investigation is prevented by the relatively few properties known up to now for the $\vartheta(\rho, V|F)$ function, and even a numerical study does not seem to be simple in the general case (see Remark 6 below). In this section, we provide a partial theoretical investigation which is enough to compare our bound with all previous versions of the Elias bound and with Lovász' and Marton's bound on graph capacity.

A. Binary Channels

We first give evidence that the proposed bound is a generalization of the Elias bound by showing in detail how the original one for binary channels is recovered as a special case. This shows that, even in the binary case, there is no loss in the use of equation (55) with the approach based on ϑ with respect to the standard use of the Plotkin bound (21) under composition constraints. In particular, the original bound for binary channels is obtained in the limit $\rho \rightarrow \infty$.

Consider a binary alphabet $\mathcal{X} = \{0, 1\}$ and distance $d(0, 1) = 1$. Then, for any ρ it is not difficult to see that one can always take as an optimal representation of degree ρ the two-dimensional vectors

$$\begin{aligned} u_0 &= [\cos(\alpha), \sin(\alpha)]^\dagger \\ u_1 &= [\cos(\alpha), -\sin(\alpha)]^\dagger \end{aligned}$$

where α satisfies $\cos(2\alpha) = e^{-1/\rho}$. For a given distribution Q , let the optimal handle which achieves $\vartheta(\rho, Q)$ be

$$f = [\cos(\beta), \sin(\beta)]^\dagger.$$

Then

$$\vartheta(\rho, Q) = -2Q(0) \log \cos(\alpha - \beta) - 2Q(1) \log \cos(\alpha + \beta). \quad (93)$$

where the value of β can be determined by minimizing this expression. Upon differentiation and a little of algebra we find

$$\sin(2\beta) = (Q(0) - Q(1)) \sin(2\alpha). \quad (94)$$

The value of $\vartheta(\rho, Q)$ can now be computed analytically by using this relation in (93). The resulting expression is complicated and not very useful here. So, we only study the bound of Theorem 6 asymptotically obtained by letting $\rho \rightarrow \infty$ with appropriate choices of F and V . We also only study the bound obtained for the uniform composition P , since we already know that this is the interesting case for the original Elias bound (see Lemma 4 below for details).

First note that, for any V , $\vartheta(\rho, V|F) \rightarrow 0$ as $\rho \rightarrow \infty$, which means that we can obtain a bound for any R by choosing F and V such that $I(F, V) < R$. Let us then choose $\mathcal{A} = \{0, 1\}$, F be uniform, and V such that $V_0(1) = V_1(0) = \lambda$, with λ such that $I(F, V) = 1 - h(\lambda) < R$, where $h(\cdot)$ is the binary entropy function. If we set $Q = V_0$, then by symmetry we have $\vartheta(\rho, V|F) = \vartheta(\rho, Q)$. Since $\cos(2\alpha) = e^{-1/\rho}$, in the limit $\rho \rightarrow \infty$ we have $\alpha \rightarrow 0$, and from equation (94) we deduce that $\beta \approx \alpha(1 - 2\lambda)$. The expression for $\vartheta(\rho, Q)$ is then asymptotically

$$\begin{aligned} \vartheta(\rho, Q) &\approx -2(1 - \lambda) \log \cos(2\lambda\alpha) - 2\lambda \log \cos(2(1 - \lambda)\alpha) \\ &\approx (1 - \lambda)(4\lambda^2\alpha^2) + \lambda(4(1 - \lambda)^2\alpha^2) \\ &= 4\lambda(1 - \lambda)\alpha^2. \end{aligned}$$

Using again the relation $e^{-1/\rho} = \cos(2\alpha)$ we deduce that

$$\rho = -\frac{1}{\log \cos(2\alpha)} \quad (95)$$

$$\approx \frac{1}{2\alpha^2}. \quad (96)$$

So, $\rho\vartheta(\rho, Q) \approx 2\lambda(1 - \lambda)$. The bound of Theorem 6 states that for $R > \vartheta(\rho, V|F) + I(F, V)$ we have $\delta^*(R, P) \leq \rho\vartheta(\rho, V|F)$. Since here $\vartheta(\rho, V|F) = \vartheta(\rho, Q) \rightarrow 0$ as $\rho \rightarrow \infty$, in this limit the theorem says that if $R > 1 - h(\lambda)$ then $\delta^*(R) \leq 2\lambda(1 - \lambda)$. This is precisely the Elias bound. One may wonder whether for finite ρ a better bound can be obtained. Unfortunately, a rigorous analysis seems to be painful, but numerical evaluation shows that this is not the case, the optimal bound is achieved as $\rho \rightarrow \infty$. Analogously, different choices of \mathcal{A} , F and V also do not improve the bound.

B. Squared Euclidean Distances

The analysis made above for binary channels can be extended to the general case of any discrete set \mathcal{X} when the distance d is a squared euclidean distance.

Assume then that $d(x, x')$ is a squared euclidean distance. According to Lemma 1, the matrix $G(\rho)$ with entries $g(x, x')^{1/\rho} = e^{-d(x, x')/\rho}$ is positive semidefinite for all ρ . Hence, there exist vectors $\{u_x\}$ such that $e^{-d(x, x')/\rho} = u_x^\dagger u_{x'}$ for all x, x' and, in particular, these $\{u_x\}$ vectors have unit norm. Hence, the set $\Gamma(\rho)$ defined in equation (56) always contains some representations that satisfy all the constraints with equality. We will consider the bound obtained for one such representation and we will focus in particular on the bound obtained as $\rho \rightarrow \infty$.

To make the following derivation easier to follow, we note first that, as $\rho \rightarrow \infty$, $u_x^\dagger u_{x'} \rightarrow 1$ for all x, x' , which means that all the vectors tend to concentrate in a very small cap on the unit sphere. Moreover, using the cosine law, as $\rho \rightarrow \infty$

$$\|u_x - u_{x'}\|^2 = 1 + 1 - 2u_x^\dagger u_{x'} \quad (97)$$

$$= 2(1 - e^{-d(x, x')/\rho}) \quad (98)$$

$$\approx \frac{2}{\rho} d(x, x'). \quad (99)$$

So, the vectors u_x tend to concentrate on a small cap and they tend to reproduce a scaled version of the original constellation of the given points with their squared distances. We will exploit this fact to show that in the limit $\rho \rightarrow \infty$ our bound has a very simple geometric interpretation, which will also allow us to connect our bound to the other ones mentioned before.

For any F and V , it is not difficult to see that the optimal choice of the handles $\{f_a\}$ in equation (81) will be such that $\vartheta(\rho, V|F) \rightarrow 0$ as $\rho \rightarrow \infty$. Hence, in the limit of $\rho \rightarrow \infty$, the bound of Theorem 6 says that if $R > I(F, V)$ then

$$\delta^*(R, P) \leq \lim_{\rho \rightarrow \infty} \rho \vartheta(\rho, V|F). \quad (100)$$

So, we are now interested in evaluating the above limit.

For a fixed value of a , consider the quantity $\vartheta(\rho, V_a)$ which appears in the definition (80). Let for ease of notation $Q = V_a$, so that we can focus for a moment on the evaluation of $\vartheta(\rho, Q)$ for a general Q and get rid of a . As mentioned before, we can pick a representation which satisfies $u_x^\dagger u_{x'} = e^{-d(x, x')/\rho}$, and all these vectors tend to concentrate in a small cap on the unit sphere as $\rho \rightarrow \infty$. The handle f of the representation will surely also be in this small cap and hence $f^\dagger u_x \rightarrow 1$ as $\rho \rightarrow \infty$. Let now θ_x be the angle between the handle f and the vector u_x . We have $|f^\dagger u_x|^2 = \cos^2(\theta_x)$ and, since $\theta_x \rightarrow 0$, we can use the expansion $-\log(\cos^2(t)) = t^2 + o(t^2)$, valid for $t \rightarrow 0$, to deduce that

$$\log \frac{1}{|f^\dagger u_x|^2} = \theta_x^2 + o(\theta_x^2). \quad (101)$$

On the other hand, we have $\|f - u_x\|^2 = 4 \sin^2(\theta_x/2) = \theta_x^2 + o(\theta_x^2)$, which implies that

$$\log \frac{1}{|f^\dagger u_x|^2} = \|f - u_x\|^2 + o(\|f - u_x\|^2). \quad (102)$$

The optimal choice of the handle f will thus be asymptotically such as to minimize a quantity of the form

$$\sum_x Q(x) (\|f - u_x\|^2 + o(\|f - u_x\|^2)). \quad (103)$$

If we neglect for a moment the $o(\cdot)$ term, we notice that the quantity to minimize is precisely the average distortion of a quantizer which uses f for representing the vectors $\{u_x\}$. Without constraints on f , it is well known that the choice of f which minimizes the distortion is the centroid and, hence, we expect the handle to satisfy

$$f \approx \sum_{x'} Q(x') u_{x'}. \quad (104)$$

However, f must be a unit norm vector and hence we cannot replace the approximation with equality in the above equation⁷. In order to simplify the discussion, instead of studying the performance obtained for the optimal handle, we show the results obtained for a suboptimal choice, which can however be proved to be the true asymptotically optimal performance with a more detailed analysis. So, we choose the suboptimal handle

$$f = \frac{\sum_{x'} Q(x') u_{x'}}{\|\sum_{x'} Q(x') u_{x'}\|}. \quad (105)$$

Then, for any x we have

$$\theta_x^2 \approx 2(1 - \cos(\theta_x)) \quad (106)$$

$$= 2 - 2 \frac{\sum_{x'} Q(x') u_{x'}^\dagger u_x}{\|\sum_{x'} Q(x') u_{x'}\|}. \quad (107)$$

Using now equation (101), we have

$$\sum_x Q(x) \log \frac{1}{|f^\dagger u_x|^2} \approx 2 - 2 \frac{\sum_{x',x} Q(x') Q(x) u_{x'}^\dagger u_x}{\|\sum_{x'} Q(x') u_{x'}\|} \quad (108)$$

$$= 2 - 2 \sqrt{\sum_{x',x} Q(x') Q(x) u_{x'}^\dagger u_x}. \quad (109)$$

The square root in the last expression can be approximated, as $\rho \rightarrow \infty$, as follows

$$\sqrt{\sum_{x',x} Q(x') Q(x) u_{x'}^\dagger u_x} = \sqrt{\sum_{x',x} Q(x') Q(x) e^{-d(x,x')/\rho}} \quad (110)$$

$$\approx \sqrt{\sum_{x',x} Q(x') Q(x) \left(1 - \frac{d(x,x')}{\rho}\right)} \quad (111)$$

$$= \sqrt{1 - \sum_{x',x} Q(x') Q(x) \frac{d(x,x')}{\rho}} \quad (112)$$

$$\approx 1 - \frac{1}{2\rho} \sum_{x',x} Q(x') Q(x) d(x,x'). \quad (113)$$

⁷In practice, as $\rho \rightarrow \infty$ all the vectors u_x and f can be considered asymptotically co-planar. However, an accurate estimation of $\vartheta(\rho, Q)$ requires some care in the use of equation (104).

In conclusion, we have the approximation

$$\sum_x Q(x) \log \frac{1}{|f^\dagger u_x|^2} \approx \frac{1}{\rho} \sum_{x',x} Q(x') Q(x) d(x, x'). \quad (114)$$

which implies that

$$\lim_{\rho \rightarrow \infty} \rho \vartheta(\rho, Q) \leq \sum_{x',x} Q(x') Q(x) d(x, x'). \quad (115)$$

If we now use this result for the generic term $\vartheta(\rho, V_a)$ which appears in the definition (80), we find that the right hand side of equation (100) can be bounded as

$$\lim_{\rho \rightarrow \infty} \rho \vartheta(\rho, P, V) \leq \sum_{a, x_1, x_2} F(a) V_a(x) V_a(x') d(x, x'). \quad (116)$$

So, the bound obtained as $\rho \rightarrow \infty$ can be stated as follows.

Theorem 8: For a squared euclidean distance d , for a distribution F on a set \mathcal{A} and a stochastic matrix $V : \mathcal{A} \rightarrow \mathcal{X}$ such that $FV = P$ and $R > I(F, V)$, we have the bound

$$\delta^*(R, P) \leq \sum_{a, x_1, x_2} P(x) V_a(x) V_a(x') d(x, x'). \quad (117)$$

To the best of our knowledge, this result is new. When we optimize over F and V to get the best possible bound and over P to get the best possible code, we obtain the following result.

Theorem 9: For a squared euclidean distance d , we have the bound

$$\delta^*(R) \leq \max_P \min_{F, V} \sum_{a, x, x'} F(a) V_a(x) V_a(x') d(x, x') \quad (118)$$

where the inner minimum is over the distributions F and V such that $FV = P$ and $I(F, V) \leq R$.

Remark 5: It can be observed that for the particular choice $\mathcal{A} = \mathcal{X}$ and $F = P$, the bound takes the form of the Blahut's bound which, as said in Section III-E, holds for all squared euclidean distances and not just for the Bhattacharyya distance.

Remark 6: We observe that the evaluation of (118), as well as Blahut's bound (48), is more complex than what could seem at first sight because, as we already mentioned in Section III-E, the objective function in equation (118) is concave in V and, hence, the minimization which appears there is not computationally simple. This problem is essentially the same encountered by Piret in the evaluation of his bound [4], and as he suggests, for this type of bounds it may just be preferable to guess good choices of F and V and numerically compute the resulting bound. Since the bound of Theorem 6 includes the bound of Theorem 8, the same remark applies to it.

C. Circularly Symmetric Distances

We now consider the particular case where $\mathcal{X} = \mathbb{Z}_K = \{0, 1, \dots, K-1\}$, endowed with its usual sum, and where the distance $d(x, x')$ is a function of $x - x'$. Note that since $d(x, x') = d(x', x)$ by assumption, $d(x, x')$ is actually a function of $|x - x'|$. Examples of such distances are the Hamming distance, the Lee distance, or the squared euclidean distance for a set of regularly spaced points on the unit circle. In this case we can simplify our bounds since, due to the symmetry, the uniform composition is optimal for any R as stated in the following Lemma.

Lemma 4: For a circularly symmetric distance, letting U be the uniform distribution, we have

$$\delta^*(R, P) \leq \delta^*(R, U). \quad (119)$$

Hence, $\delta^*(R) = \delta^*(R, U)$.

Proof: The proof of the Lemma is based on a constructive procedure. For a given code \mathcal{C} of length n , with $M = e^{nR}$ codewords of composition P and minimum distance $d_{\min}(\mathcal{C})$, we can construct a code $\tilde{\mathcal{C}}$ with constant composition Q such that $|Q(x) - 1/K| \leq \varepsilon_n$, minimum distance $d_{\min}(\tilde{\mathcal{C}}) \geq d_{\min}(\mathcal{C})$ and rate $\tilde{R} = R - \alpha_n$, where $\varepsilon_n \rightarrow 0$ and $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$.

Let $\mathcal{C} = \{\mathbf{x}_1, \dots, \mathbf{x}_M\}$. Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be a random sequence of uniform independent symbols from \mathcal{X} and set

$$\tilde{\mathbf{X}}_m = \mathbf{x}_m + \mathbf{X}, \quad m = 1, 2, \dots, M. \quad (120)$$

First note that $d(\tilde{\mathbf{X}}_m, \tilde{\mathbf{X}}_{m'}) = d(\mathbf{x}_m, \mathbf{x}_{m'})$ and, hence, the random code so constructed has the same minimum distance as the original code. Since \mathbf{X} is uniformly distributed over \mathcal{X}^n , $\tilde{\mathbf{X}}_m$ is also uniformly distributed over \mathcal{X}^n . Let $T(\tilde{\mathbf{X}}_m)$ be the composition of the sequence $\tilde{\mathbf{X}}_m$. Let \mathcal{U}_ε be the set of distributions Q such that $|Q(x) - 1/K| \leq \varepsilon$. Finally let ε_n be the smallest ε for which the following inequality holds

$$P[T(\tilde{\mathbf{X}}_m) \in \mathcal{U}_\varepsilon] \geq 1 - \varepsilon. \quad (121)$$

Since $\tilde{\mathbf{X}}_m$ is uniformly distributed over \mathcal{X}^n , by the strong law of large numbers, $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$. Due to equation (121), the expected number of codewords $\tilde{\mathbf{X}}_m$ whose composition is in $\mathcal{U}_{\varepsilon_n}$ is at least $(1 - \varepsilon_n)M$. This implies that there exists a sequence $\tilde{\mathbf{x}} \in \mathcal{X}^n$ such that at least $M(1 - \varepsilon_n)$ of the codewords

$$\tilde{\mathbf{x}}_m = \mathbf{x}_m + \tilde{\mathbf{x}}, \quad m = 1, 2, \dots, M, \quad (122)$$

have composition in $\mathcal{U}_{\varepsilon_n}$. Since there are less than $(n+1)^K$ different compositions in $\mathcal{U}_{\varepsilon_n}$, at least $M(1 - \varepsilon_n)/(n+1)^K$ codewords $\tilde{\mathbf{x}}_m$ have the same composition. This codewords form a constant composition code $\tilde{\mathcal{C}}$ with rate $\tilde{R} \geq R - O(\log n/n)$, with minimum distance $d_{\min}(\tilde{\mathcal{C}}) \geq d_{\min}(\mathcal{C})$ and composition in $\mathcal{U}_{\varepsilon_n}$. As $n \rightarrow \infty$, this implies that $\delta^*(R, P) \leq \delta^*(R, U)$ and thus that $\delta^*(R) = \delta^*(R, U)$. ■

We can now present the simplification of Theorem 6 for the case of circularly symmetric distances. Since the uniform composition is always optimal, we can focus on the case where P is the uniform distribution. We can also consider the particular choice $\mathcal{A} = \mathcal{X}$, F uniform on \mathcal{X} , and the matrix V to satisfy $V_x(x') = Q(x' - x)$ for some distribution Q . This implies that $I(F, V) = \log K - H(Q)$ and $\vartheta(\rho, V|F) = \vartheta(\rho, Q)$. Then from Theorem 6 we deduce the following.

Theorem 10: For a circularly symmetric distance we have the bound

$$\text{if } R > \log K - H(Q) + \vartheta(\rho, Q), \quad \text{then } \delta^*(R) \leq \rho \vartheta(\rho, Q). \quad (123)$$

Finally, we can consider the particular case where the distance is a circularly symmetric squared euclidean distance. Then we can combine the simplifications used to obtain Theorems 8 and 10 to obtain the following.

Theorem 11: For a circularly symmetric squared euclidean distance, if Q satisfies $R > \log K - H(Q)$, then

$$\delta^*(R) \leq \sum_{x, x'} Q(x)Q(x')d(x, x'). \quad (124)$$

Note that this bound is essentially the same as given by Piret for the particular case of the squared euclidean distance for regularly spaced points on the unit circle. Hence, Piret's bound is contained as a particular case of Blahut's bound, which is a special case of our own.

We finally show that, Berlekamp's bound can be interpreted as a weakened version of Theorem 11. For a given distribution Q , let

$$d(Q) = \sum_x Q(x)d(0, x), \quad (125)$$

the average distance from point 0, that we use as a reference. Consider again the convex set of distributions

$$\mathcal{Q}(R) = \{Q : R \geq \log K - H(Q)\}. \quad (126)$$

We can use in Theorem 11 any Q in $\mathcal{Q}(R)$, and we choose to use the distribution Q which minimizes $d(Q)$ over $\mathcal{Q}(R)$. Thus, let

$$t := \min_{Q \in \mathcal{Q}(R)} d(Q) \quad (127)$$

and let Q^* be a minimizing Q . Then, we have the bound

$$\delta^*(R) \leq \sum_{x, x'} Q^*(x)Q^*(x')d(x, x') \quad (128)$$

$$\leq \max_{Q: d(Q)=t} \sum_{x, x'} Q(x)Q(x')d(x, x'). \quad (129)$$

Since d is a squared euclidean distance, the quadratic form in the last expression is a concave function and, thus, the maximization can be solved by means of the usual Kuhn-Tucker conditions. It can be observed that this evaluation is the same needed in Berlekamp's procedure (cf. [3, eqs. (13.63-13.66)]). The maximizing Q is of the form

$$Q(x) = \begin{cases} \frac{t}{Kd(U)} + 1 - \frac{t}{d(U)} & \text{if } x = 0 \\ \frac{t}{Kd(U)} & \text{if } x \neq 0. \end{cases} \quad (130)$$

where U is the uniform distribution. For this Q we have

$$\sum_{x, x'} Q(x)Q(x')d(x, x') = t \left(2 - \frac{t}{d(U)} \right) \quad (131)$$

So, Theorem 11 implies that

$$\delta^*(R) \leq t \left(2 - \frac{t}{d(U)} \right). \quad (132)$$

where t is defined in equation (127). This is in fact Berlekamp's extension of the Elias bound [3, Th. 13.67]. In conclusion, we have shown that our bound includes Blahut's bound as a particular case, which in turns includes Piret's which finally implies Berlekamp's one.

D. Infinite Distances: a Critical Look

In the case of infinite distances Theorem 6 exhibits both interesting properties as well as clear weaknesses. In this case, we know that even bounds on $R^*(\infty, P)$ and $R^*(\infty)$ (or on $C(G, P)$ and $C(G)$, in the graph theory language) are hard, and we first remind that, as mentioned in Remark 4, Lovász' and Marton's bounds on the capacity of graphs are recovered from Theorem 6 with a trivial choice of \mathcal{A} , F and V . The performance of the bound for $R > R(\infty)$, however, strongly depends on the particular type of graph G which is induced by finite values of d on $\mathcal{X} \times \mathcal{X}$, and it is certainly not yet satisfactory in the general case. A general analysis is prohibitively complex, but it will be useful to consider two particular cases for which we can perform simple sanity checks on our bound (see Figure 2).

Let $\mathcal{X} = \{0, 1, 2, 3\}$ be the vertices of a square and let $d(x, x') = 1$ if x, x' are adjacent, while $d(x, x') = \infty$ otherwise (see Figure 2(a)). The graph induced by finite distances is the square itself and its capacity is $\log(2)$. Due to symmetry, we test the simplified version of the bound given in Theorem 10. If we choose

$$Q(x) = \begin{cases} 1 - \lambda & \text{if } x = 0 \\ \lambda & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases} \quad (133)$$

and consider the result obtained as $\rho \rightarrow \infty$, the bound reduces to the statement that for $R > \log(4) - h(\lambda)$, $\delta^*(R) \leq 2\lambda(1 - \lambda)$. Note that the bound on $\delta^*(R)$ is smaller than $1/2$ at all rates at which it is bounded, that is for $R > \log(2)$ (the capacity of the graph, indeed). In fact, it can be observed that the bound is exactly the standard Elias bound for binary codes shifted by a quantity $\log(2)$ on the R axis. This is in accordance with intuition, since it is not difficult to see that at rates $R > \log(2)$ there are at least $e^{n(R - \log(2))}$ codewords which are all at finite distance and which can be mapped to a binary alphabet without modifying the distances among them. Thus, bounds on $\delta^*(R)$ for the original setting can be deduced from bounds on $\delta^*(R - \log(2))$ for binary codes. This is automatically taken care of in our bound and, hence, in this case we can say that the bound is a satisfactory extension of the standard bound. It is not difficult to see that this happens for all even cycles.

Consider instead the case of the pentagon with vertex set $\mathcal{X} = \{0, 1, 2, 3, 4\}$, and where we let $d(x, x') = 1$ if x, x' are adjacent in the pentagon, while $d(x, x') = \infty$ otherwise (see Figure 2(b)). The graph induced by finite distances is the pentagon itself and its capacity is $\log(5)/2$ [5]. If we apply Theorem 10 with the same choice of Q mentioned above and letting $\rho \rightarrow \infty$, we get the bound $\delta^*(R) \leq 2\lambda(1 - \lambda)$ for $R > \log(5) - h(\lambda)$. So, this choice of Q only gives a finite (and reasonably good) bound for $R > \log(5/2) > \log(5)/2$. This was to be expected, since we are essentially not using the Lovász theta function of the pentagon⁸ with this choice of Q . So, finite bounds on $\delta^*(R)$ for $\log(5)/2 < R < \log(5/2)$ require, as is obvious, other choices of Q and ρ . A detailed analysis is complicated and, as mentioned, the evaluation of the bound is not simple in general and will be hopefully

⁸We are using a Q which is good enough to obtain the fractional clique covering number of the graph. Actually, there is even no need to use ϑ functions for this choice of Q , since we reduce the problem to binary alphabets. We do not go into these details and leave further analysis to future works.

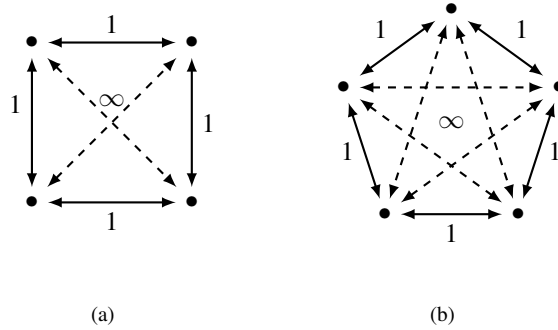


Fig. 2. The two examples of distances discussed in Section V-D.

investigated in more detail in a future work. We can here at least mention that, to the best of our understanding, for rates slightly larger than $\log(5)/2$, no choices of ρ and Q lead to a bound on $\delta^*(R)$ which is as good as the trivial bound $\delta^*(R) \leq 1$. This is of course a frustrating negative point on our bound. One may ask whether other choices of \mathcal{A} , F and V in the original bound of Theorem 6 would give better results than Theorem 10. We tend to exclude this, though we do not have a rigorous prove.

The pentagonal example discussed above shows that, even if the bound has the reasonably good property of including all previous versions of the Elias bound as well as Lovász' and Marton's bounds on graph capacities, it is still surely not a satisfactory bound in the case of general possibly infinite valued distances. We propose the study of bounds on $\delta^*(R)$ for the pentagonal example proposed here as an interesting open problem which deserves further attention.

VI. RELIABILITY FUNCTION

We present here an important case of application of the bound, that is, its use in bounding the reliability function of classical and classical-quantum channels. We describe this two cases separately for the reader's convenience.

A. Classical Channels

Let $\mathcal{X} = \{1, 2, \dots, |\mathcal{X}|\}$ and $\mathcal{Y} = \{1, 2, \dots, |\mathcal{Y}|\}$ be the input and output alphabets of a discrete memoryless channel with transition probabilities $W_x(y)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. If $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is a sequence of n input symbols and correspondingly $\mathbf{y} = (y_1, y_2, \dots, y_n)$ is a sequence of output symbols, then the probability of observing \mathbf{y} at the output of the channel given input \mathbf{x} is

$$W_{\mathbf{x}}(\mathbf{y}) = \prod_{i=1}^n W_{x_i}(y_i).$$

An (M, n) code is a set of M n -symbol sequences $\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$, $\mathbf{x}_m \in \mathcal{X}^n$ associated to M messages $\{1, \dots, M\}$, and a decoder is a map from the set of output sequences \mathcal{Y}^n to $\{1, 2, \dots, M\}$. Let \mathcal{Y}_m be the set of output sequences

that are mapped to the message m by the decoder. When message m is sent, the probability of error is

$$P_{e|m} = 1 - \sum_{y \in \mathcal{Y}_m} W_{x_m}(y).$$

The maximum error probability of the code is defined as the largest $P_{e|m}$, that is,

$$P_{e,\max} = \max_m P_{e|m}.$$

Let $P_{e,\max}^{(n)}(R)$ be the smallest maximum error probability among all codes of length n and rate at least R . The reliability function is defined as (see [16] for more details)

$$E(R) = \limsup_{n \rightarrow \infty} -\frac{1}{n} \log P_{e,\max}^{(n)}(R). \quad (134)$$

For any channel, for a given code, the probability of error $P_{e,\max}$ is lower bounded by the probability of error in any binary hypothesis test between two codewords. In a binary hypothesis test between codewords m and m' , an extension of the Chernoff Bound allows to assert that the minimum error probability P_e vanishes exponentially fast in the block length n and that [17]

$$\log \frac{1}{P_e} \leq D_C(\mathbf{W}_{x_m}, \mathbf{W}_{x_{m'}}) + o(n).$$

where $D_C(\cdot, \cdot)$ is the Chernoff distance between two distributions defined by

$$D_C(Q_1, Q_2) = -\log \inf_{0 < s < 1} \sum_y Q_1(y)^{1-s} Q_2(y)^s. \quad (135)$$

Note that we use a different notation for the Chernoff distance because it is not additive, in the sense that, in general,

$$D_C(\mathbf{W}_{x_m}, \mathbf{W}_{x_{m'}}) \neq \sum_{i=1}^n D_C(W_{x_{m,i}}, W_{x_{m',i}}). \quad (136)$$

Using the above considerations, $E(R)$ can be bounded as

$$E(R) \leq \frac{1}{n} \min_{m \neq m'} D_C(\mathbf{W}_{x_m}, \mathbf{W}_{x_{m'}}) + o(1). \quad (137)$$

Hence, upper bounds on $E(R)$ can be deduced by determining upper bounds on the minimum Chernoff distance which appears on the right hand side of (137). Due to equation (136), we cannot apply our bound on the minimum distance directly to the Chernoff distance, but we can use additive distances which upper bound it. The Bhattacharyya distance can be used for this purpose; it can be proved that

$$d_B(x, x') \leq D_C(W_x, W_{x'}) \leq 2d_B(x, x') \quad (138)$$

For the so called *pairwise reversible channels* [17], we have equality on the left hand side for all x, x' and hence

$$E(R) \leq \delta_B^*(R). \quad (139)$$

Thus, our bounds on $\delta_B^*(R)$ apply directly to $E(R)$. However, for other channels, equality holds on the right hand side of (138). For these channels, the best that we can do in bounding $E(R)$ using d_B is using the inequality

$$E(R) \leq 2\delta_B^*(R). \quad (140)$$

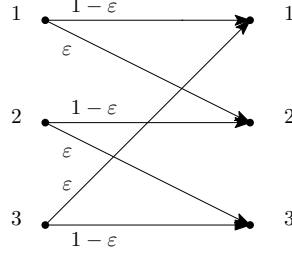


Fig. 3. The “ternary unilateral channel” of [17, Fig. 3].

We observe in particular that Blahut’s proof of his upper bound on $E(R)$ in [2, Th. 12] only holds for pairwise reversible channels. The problem for general channels comes from [2, Sec. VI, page 669, second column] where it is stated that “Now $x_{m'}$ and x_m have the same composition [...] ; hence, the first term is zero”. This statement is not correct, since it would essentially imply that the Chernoff distance between two codewords with the same composition equals their Bhattacharyya distance, which is not always the case. More specifically, in our notation, the quoted statement is that if x and x' are two codewords with the same composition, then, setting

$$Q(y) = \frac{\sqrt{W_x(y)W_{x'}(y)}}{\sum_{y'} \sqrt{W_x(y')W_{x'}(y')}} \quad (141)$$

we have

$$\sum_y Q(y) \log \frac{W_x(y)}{W_{x'}(y)} = 0. \quad (142)$$

This is not true, as proved by the “ternary unilateral channel” mentioned in [17] and shown in Fig. 3. For the codewords $x = (1, 2, 3)$ and $x' = (2, 3, 1)$, which have the same composition, we have

$$Q(y) = \begin{cases} 1 & \text{if } y = (2, 3, 1) \\ 0 & \text{otherwise} \end{cases} \quad (143)$$

and consequently

$$\sum_y Q(y) \log \frac{W_x(y)}{W_{x'}(y)} = 3 \log \frac{1 - \varepsilon}{\varepsilon}. \quad (144)$$

One of the consequences of this observation is that, contrarily to what stated in [2, Sec. I], Blahut’s bound does not include the zero-rate bound of [17] for channels that are not pairwise reversible. So, the only proof that the expurgated bound is tight at $R = 0$ for any channel without a zero-error capacity remains Berlekamp’s complicated proof [18], [17].

For channels that are not pairwise reversible but for which $D_C(W_x, W_{x'})$ is only slightly larger than $d_B(x, x')$, using equation (140) can be very suboptimal. A possible alternative approach is to introduce an additive variation of the Chernoff distance. We can define

$$d_{\tilde{C}}(x, x') = D_C(W_x, W_{x'}) \quad (145)$$

and extend $d_{\tilde{C}}$ additively to sequences. It is then not difficult to see that for sequences \mathbf{x}, \mathbf{x}'

$$d_{\tilde{C}}(\mathbf{x}, \mathbf{x}') \geq D_C(\mathbf{W}_{\mathbf{x}_m}, \mathbf{W}_{\mathbf{x}_{m'}}). \quad (146)$$

So, we have

$$E(R) \leq \delta_{\tilde{C}}^*(R) \quad (147)$$

and we can thus bound $E(R)$ by using our bound for the distance $d_{\tilde{C}}$. Note in particular that this recovers the particular case of pairwise reversible channels, since $d_{\tilde{C}} = d_B$ in that case. On the other hand, for the channel of Fig. 3, as $\varepsilon \rightarrow 0$ $d_{\tilde{C}}(x, x')/d_B(x, x') \rightarrow 2$, and we thus recover equation (140) which, as $R \rightarrow 0$, is loose by a factor of two (cf. [17]).

B. Classical-Quantum Channels

Consider a classical-quantum channel with input alphabet $\mathcal{X} = \{1, \dots, |\mathcal{X}|\}$ and associated density operators S_x , $x \in \mathcal{X}$, in a finite dimensional Hilbert space \mathcal{H} . The n -fold product channel acts in the tensor product space $\mathcal{H} = \mathcal{H}^{\otimes n}$ of n copies of \mathcal{H} . To a sequence $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is associated the signal state $\mathbf{S}_{\mathbf{x}} = S_{x_1} \otimes S_{x_2} \cdots \otimes S_{x_n}$. As in the classical case, an (M, n) code is a set of M n -symbol sequences $\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$, $\mathbf{x}_m \in \mathcal{X}^n$ associated to M messages $\{1, \dots, M\}$. A quantum decision scheme for such a code is a so-called POVM (see for example [19]), that is, a collection of M positive operators $\{\Pi_1, \Pi_2, \dots, \Pi_M\}$ such that $\sum \Pi_m \leq \mathbb{1}$, where $\mathbb{1}$ is the identity operator. The probability that message m' is decoded when message m is transmitted is $P_{m'|m} = \text{Tr} \Pi_{m'} \mathbf{S}_{\mathbf{x}_m}$. The probability of error after sending message m is

$$P_{e|m} = 1 - \text{Tr} (\Pi_m \mathbf{S}_{\mathbf{x}_m}).$$

We then define $P_{e,\max}$, $P_{e,\max}^{(n)}(R)$ and $E(R)$ precisely as in the classical case.

With the same reasoning used for classical channels (see [8]) we come to the conclusion that

$$E(R) \leq \frac{1}{n} \min_{m \neq m'} D_C(\mathbf{S}_{\mathbf{x}_m}, \mathbf{S}_{\mathbf{x}_{m'}}) + o(1), \quad (148)$$

where $D_C(\cdot, \cdot)$ is now the Chernoff distance between two density operators

$$D_C(A, B) = -\log \inf_{0 < s < 1} \text{Tr} A^{1-s} B^s. \quad (149)$$

Again we can use bounds on D_C based on additive distances to bound $E(R)$. In particular, we can use the Bhattacharyya distance

$$d_B(x, x') = -\log \text{Tr} S_x^{1/2} S_{x'}^{1/2}, \quad (150)$$

for which we have

$$d_B(x, x') \leq D_C(S_x, S_{x'}) \leq 2d_B(x, x'), \quad (151)$$

with equality again on the left for pairwise reversible channels. In the quantum setting, a particularly importance case is given by pure-state channels with states $S_x = |\psi_x\rangle\langle\psi_x|$, for which we always have $d_B(x, x') = D_C(S_x, S_{x'}) =$

$|\langle \psi_x | \psi_{x'} \rangle|^2$. For classical-quantum channels we can also use the bound

$$D_C(S_x, S_{x'}) \leq 2d_F(x, x'), \quad (152)$$

where

$$\begin{aligned} d_F(A, B) &= -\log \text{Tr} |\sqrt{A}\sqrt{B}| \\ &= -\log \text{Tr} \sqrt{\sqrt{A} B \sqrt{A}}. \end{aligned}$$

We still come to the conclusion, however, that the best choice is simply to use an additive variation of the Chernoff distance

$$d_{\tilde{C}}(x, x') = D_C(S_x, S_{x'}), \quad (153)$$

as for the classical case.

VII. ACKNOWLEDGMENTS

The author would like to thank Telecom Italia Lab and, in particular, Skjalg Lepsøy and Gianluca Francini for introducing him to euclidean embedding during the 2008 project DYNAMIC TV. Useful discussions with Richard Blahut and Yury Polyanskiy are also acknowledged.

APPENDIX

PROOF OF LEMMA 1

The core part of Lemma 1 is by now a classic result in the theory of positive definite kernels and functions, and should really be interpreted in that context. A detailed discussion can be found for example in [20]. Only condition (c), which is much important for us, is apparently not usually mentioned in that context. In the context of information theory, Jelinek already used the central part of the lemma in his paper [13]. Since we need in any case to add some integration to those references, we provide a complete self contained proof for the reader convenience.

Proof of Lemma 1: We break down the proof into single implications which, altogether, imply the lemma.

- Implication (a) \Rightarrow (b).

This is a known connection between infinitely divisible and negative almost definite kernels. Assume $G(\rho)$ is positive semidefinite for all $\rho > 0$ and that $\sum_x c(x) = 0$. Then

$$\begin{aligned} \sum_{x, x'} d(x, x') c(x) c(x') &= \lim_{\rho \rightarrow \infty} \sum_{x, x'} \rho \left(1 - e^{-d(x, x')/\rho} \right) c(x) c(x') \\ &= -\rho \sum_{x, x'} e^{-d(x, x')/\rho} c(x) c(x') \\ &\leq 0, \end{aligned}$$

where we have used the fact that $\sum_{x, x'} c(x) c(x') = 0$.

- Implications (b) \Leftrightarrow (c).

We introduce some notation here and prove a slightly stronger result which will also be useful in the next step of the proof. For any two functions q_1 and q_2 on \mathcal{X} , let

$$f(q_1, q_2) := \sum_{x, x'} q_1(x) q_2(x') d(x, x'), \quad f(q) := f(q, q). \quad (154)$$

We prove that $f(q)$ is concave on every affine hyperplane defined by $\sum_x q(x) = t$, with t a constant, if and only if $f(q) \leq 0$ whenever $\sum_x q(x) = 0$. Observe that we need only prove midpoint concavity here (a general proof is not substantially different, but requires a more complicated notation). Note that for two functions a and b on \mathcal{X} , due to the symmetry of d we have

$$f(a + b) = f(a) + f(b) + 2f(a, b) \quad (155)$$

and, hence

$$\frac{f(a + b) + f(a - b)}{2} = f(a) + f(b) \quad (156)$$

Now, assume $\sum_x q_1(x) = \sum_x q_2(x) = t$, and define the functions $q = (q_1 + q_2)/2$ and $c = (q_1 - q_2)/2$. Note that $\sum_x q(x) = t$ and $\sum_x c(x) = 0$. Then,

$$\begin{aligned} \frac{f(q_1) + f(q_2)}{2} &= \frac{f(q + c) + f(q - c)}{2} \\ &= f(q) + f(c) \\ &= f\left(\frac{q_1 + q_2}{2}\right) + f(c). \end{aligned}$$

So, if $f(c) \leq 0$ then f is midpoint concave (and hence concave) on the affine hyperplane defined by $\sum_x q(x) = t$. Since we used a one to one map $(q_1, q_2) \leftrightarrow (q, c)$, we can invert the reasoning and find, for any c , an appropriate pair q_1 and q_2 to show that if f is concave in any such hyperplane, then $f(c) \leq 0$ whenever $\sum_x c(x) = 0$.

- Implication (b) \Rightarrow (d)

Define

$$s(x) := \frac{1}{|\mathcal{X}|} \sum_{x'} d(x, x'), \quad s := \frac{1}{|\mathcal{X}|^2} \sum_{x, x'} d(x, x'), \quad (157)$$

then

$$\tilde{d}(x, x') := -d(x, x') + s(x) + s(x') - s \quad (158)$$

and finally, for functions a and b on \mathcal{X} ,

$$\tilde{f}(a, b) := \sum_{x, x'} a(x) b(x') \tilde{d}(x, x'), \quad \tilde{f}(a) := f(a, a). \quad (159)$$

Again, by symmetry of $\tilde{d}(x, x')$ we have

$$\tilde{f}(a + b) = \tilde{f}(a) + 2\tilde{f}(a, b) + \tilde{f}(b). \quad (160)$$

Furthermore, it is not difficult to see that if $\sum_x c(x) = 0$, then $\tilde{f}(c) = -f(c)$, where f is defined as in (154). In addition, a direct calculation shows that if b is constant, then $\tilde{f}(a, b) = 0$ for any a , which implies that $\tilde{f}(a + b) = \tilde{f}(a) + 2\tilde{f}(a, b) + \tilde{f}(b) = \tilde{f}(a)$.

Now, for any a , choose $b(x) = \sum_{x'} a(x')/|\mathcal{X}|$, so that b is constant and $\sum_x (a(x) - b(x)) = 0$. Then, using the properties mentioned above we have

$$\tilde{f}(a) = \tilde{f}(a - b) \quad (161)$$

$$= -f(a - b) \quad (162)$$

$$\geq 0, \quad (163)$$

where in the last step we have used the condition (b) of the Lemma with the choice $c(x) = a(x) - b(x)$.

So, the matrix \tilde{D} with elements $\tilde{d}(x, x')$ is positive semidefinite and, hence, it is a Gram matrix, which means that there exists a set of vectors $\{v_x\}_{x \in \mathcal{X}}$ such that $\tilde{d}(x, x') = v_x^\dagger v_{x'}$. Then, we have

$$\|v_x - v_{x'}\|^2 = v_x^\dagger v_x - 2v_x^\dagger v_{x'} + v_{x'}^\dagger v_{x'} \quad (164)$$

$$= \tilde{d}(x, x) - 2\tilde{d}(x, x') + \tilde{d}(x', x') \quad (165)$$

$$= -d(x, x) - d(x', x') + 2d(x, x') \quad (166)$$

$$= 2d(x, x'), \quad (167)$$

where we have used equation (158) and the fact that $d(x, x) = 0$ for all x by assumption. Hence, $d(x, x') = \|u_x - u_{x'}\|^2$ if we set $u_x = v_x/\sqrt{2}$, that is, d is a squared euclidean distance.

- Implication (d) \Rightarrow (a)

This is by now a well known basic property extensively used in the theory of reproducing kernel Hilbert spaces. We need only prove the implication for $\rho = 1$, since $d(x, x')/\rho$ is a squared euclidean distance whenever $d(x, x')$ is. Then, for any $a(x)$ we have

$$\sum_{x, x'} a(x)a(x')e^{-d(x, x')} = \sum_{x, x'} a(x)a(x')e^{-\|v_x - v_{x'}\|^2} \quad (168)$$

$$= \sum_{x, x'} \frac{a(x)}{e^{\|v_x\|^2}} \frac{a(x')}{e^{\|v_{x'}\|^2}} e^{2v_x^\dagger v_{x'}} \quad (169)$$

$$= \sum_{x, x'} b(x)b(x') \sum_{k=0}^{\infty} \frac{(2v_x^\dagger v_{x'})^k}{k!} \quad (170)$$

where $b(x) = a(x)/\exp(\|v_x\|^2)$. Denoting with \cdot^{\otimes} the k -fold Kronecker power of a vector, we can then rewrite the last expression to get

$$\begin{aligned} \sum_{x, x'} a(x)a(x')e^{-d(x, x')} &= \sum_{k=0}^{\infty} \frac{2^k}{k!} \sum_{x, x'} (b(x)v_x^{\otimes k})^\dagger (b(x')v_{x'}^{\otimes k}) \\ &= \sum_{k=0}^{\infty} \frac{2^k}{k!} \left\| \sum_x b(x)v_x^{\otimes k} \right\|^2 \\ &\geq 0. \end{aligned}$$

Hence the matrix with elements $e^{-d(x, x')}$ is positive semidefinite, and this concludes the proof of the lemma. \blacksquare

REFERENCES

- [1] M. Dalai, “An Elias Bound on the Bhattacharyya Distance of Codes for Channels with a Zero-Error Capacity,” in *Proc. IEEE Intern. Symp. Inform. Theory*, 2014.
- [2] R. Blahut, “Composition bounds for channel block codes,” *IEEE Trans. Inform. Theory*, vol. 23, no. 6, pp. 656 – 674, nov 1977.
- [3] E. Berlekamp, *Algebraic Coding Theory*, ser. McGraw-Hill series in systems science. Aegean Park Press, 1984.
- [4] P. Piret, “Bounds for Codes Over the Unit Circle,” *Information Theory, IEEE Transactions on*, vol. 32, no. 6, pp. 760–767, 1986.
- [5] L. Lovász, “On the Shannon Capacity of a Graph,” *IEEE Trans. Inform. Theory*, vol. 25, no. 1, pp. 1–7, 1979.
- [6] K. Marton, “On the Shannon Capacity of Probabilistic Graphs,” *Journal of Combinatorial Theory, Series B*, vol. 57, no. 2, pp. 183 – 195, 1993.
- [7] M. Dalai, “An “Umbrella” Bound of the Lovász-Gallager Type,” in *Proc. IEEE Intern. Symp. Inform. Theory*, 2013, pp. 3025–3029.
- [8] —, “Lower Bounds on the Probability of Error for Classical and Classical-Quantum Channels,” *IEEE Trans. Inform. Theory*, vol. 59, no. 12, pp. 8027 – 8056, 2013.
- [9] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.
- [10] J. Korner and A. Orlitsky, “Zero-error information theory,” *IEEE Trans. on Inform. Theory*, vol. 44, no. 6, pp. 2207 –2229, Oct. 1998.
- [11] I. Csiszár and J. Körner, “On the Capacity of the Arbitrarily Varying Channel for Maximum Probability of Error,” *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 57, no. 1, pp. 87–101, 1981.
- [12] C. E. Shannon, “The Zero-Error Capacity of a Noisy Channel,” *IRE Trans. Inform. Theory*, vol. IT-2, pp. 8–19, 1956.
- [13] F. Jelinek, “Evaluation of Expurgated Error Bounds,” *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 501–505, 1968.
- [14] I. J. Schoenberg, “On Certain Metric Spaces Arising From Euclidean Spaces by a Change of Metric and Their Imbedding in Hilbert Space,” *Annals of Mathematics*, vol. 38, no. 4, pp. pp. 787–793, 1937.
- [15] M. Dalai and A. Winter, “Constant Compositions in the Sphere Packing Bound for Classical-Quantum Channels,” in *Proc. IEEE Intern. Symp. Inform. Theory*, 2014.
- [16] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, “Lower Bounds to Error Probability for Coding in Discrete Memoryless Channels. I,” *Information and Control*, vol. 10, pp. 65–103, 1967.
- [17] —, “Lower Bounds to Error Probability for Coding in Discrete Memoryless Channels. II,” *Information and Control*, vol. 10, pp. 522–552, 1967.
- [18] E. R. Berlekamp, “Block coding with noiseless feedback,” Ph.D. dissertation, MIT, Cambridge, MA, 1964.
- [19] M. M. Wilde, *Quantum Information Theory*. Cambridge University Press, 2013. [Online]. Available: <http://dx.doi.org/10.1017/CBO9781139525343>
- [20] C. Berg, J. P. R. Christensen, and P. Ressel, *Harmonic Analysis on Semigroups*. Springer-Verlag, 1984.